1998

An Example of Model-Based Empirical Research: A Soup-To-Nuts Evaluation of Alternative C2 Architectures

Benson, R.

An Example of Model-Based Empirical Research: A Soup-To-Nuts Evaluation of Alternative C2 Architectures

An Example of Model-Based Empirical Research: A Soup-To-Nuts Evaluation of Alternative C2 Architectures

R. Benson[1], W. Kemple[1], D. Kleinman[1], G. Porter[1] and D. Serfaty[2*]

1C4I Academic Group, Naval Postgraduate School, 589 Dyer Rd., Monterey, CA 93943

Corresponding author: David Kleinman (kleinman@nps.navy.mil)

2APTIMA, Inc., Burlington, MA

Abstract

This paper will describe the third in a series of experiments conducted at the Naval Postgraduate School in November 1997. In particular, we will discuss the activities and results that led to form the hypotheses for test in the experiment, and the efforts that were needed to actually run the experiment: scenario design, pre-experiment modeling, training and experimental design, data collection methods and instruments, simulator software, etc. The detailed analyses of the experimental results and findings may be found in companion papers in these proceedings.

The Adaptive Architectures in Command and Control (A2C2) project is an ambitious ONR-sponsored research initiative to: extend 12 years of naval composite warfare decisionmaking research into the joint C2 arena; focus on adaptive architectures within these decision-making organizations; and produce results ranging from purely theoretical to those that can be used by the operational forces in the near term. This "industry-university-government" initiative involves a three-pronged, coordinated research effort that includes field, experimental, and theoretical components. The theoretical and analytical efforts provide models of decision-making and adaptation that are tested via experiments with military officers in joint settings for measurement of individual and team performance in dynamically evolving missions and scenarios.

1. **The Model-Test-Model Paradigm**

Our iterative approach combining the theoretical and experimental activities in the A2C2 project is based on a "model-test-model paradigm". Hypotheses, generated from either model predictions (or findings/results from previous experiments), are tested via team-in-the-loop experiments. The results from these experiments are compared with the *a priori* predictions of performance and/or process measures as generated via the models. When these alternative sources of data are not in agreement (as is often the case), we re-examine our models - as well as our experimental procedures - to explain differences and discrepancies. Our models are generally modified by including sub-models, constraints and/or biases, and subjective factors pertaining to the human decisionmaking processes that may not have been part of the original model. The subsequent model(s) that emerge from this process are then again used to predict human/team performance and process measures in a different but related context, and the process repeats. This type of pragmatic approach to model development assures that model predictions are brought into agreement with data through causal and explainable mechanisms.

For the A2C2 project, the results from experiment 2 provided a basis for the hypotheses to be examined in experiment 3. Researchers adjusted the structure of their analytical models based on findings that emerged from experiment 2. In particular, teams in experiment 2 did not have the opportunity to alter the architecture in which they performed. In the data collection phase, they only provided "pen and paper" responses on how the current architecture could be modified to perform the mission more efficiently. Experiment 3 went a step further by providing teams with *predetermined* alternative architectures, each with different advantages and disadvantages, that the teams could choose to operate under. They were then required to play the simulation again under their new architecture of choice and one of the other post-trigger architectures.

1.1 **Findings from Experiment 2**

The second experiment performed under the A2C2 project was described and presented at last year's C2 symposium [Kemple, et al., 1996]. The effort involved a (simulated) amphibious mission requiring subjects to land troops at two beaches, proceed inland along roads, and seize an airport and seaport. The enemy reacted with a host of hostile air, sea and land assets that the subjects needed to defend their own assets and bases against. Two different six-node organizational structures were designed, and served as the two levels of the independent variable 'architecture'. A model-based architecture, developed at the University of Connecticut (UConn), was designed to allocate resources, responsibility and authority in such a manner as to optimize (minimize) workload associated with inter-nodal coordination {Levchuck, et al. 1997]. The second architecture was developed by Lead Team students at the Naval Postgraduate School (NPS), and was based on application of traditional military design principles [Kemple, et al., 1997]. The NPS lead team consisted of mid-level officers (O3 to O4) in the Joint Command, Control, Communications, Computers, and Intelligence (JC4I) curriculum. These officers, representative of each service, were considered subject matter experts and performed various roles in support of the experiment. A total of six 6-person teams from the JC4I junior class and the Space Systems Operations curriculum provided the subject pool for the experimental runs. Three of these teams played in the 'optimal' design, and the three other teams played in the 'traditional' architecture.

Besides the comparisons of performance and processes between teams in each architecture, the <u>overall</u> effort attempted to explore adaptation/response of each architecture to an external 'trigger' event. The event presented to the teams was a significant (~30%) reduction in own forces, while still requiring the same mission to be completed. Each team was asked, via an *off-line*, post-experiment planning session, to produce an organization with detailed command structure (who reports to whom), resource structure (who owns what), communication structure (who communicates with whom), etc. The elicited organizational designs were not actually played out, but were intended to serve as inputs to researchers for the design of the next experiment.

The findings with respect to team adaptation generally showed that teams in the traditional architecture elected to stay within a traditional command structure with minimal reallocation of assets in the post-trigger design. The teams in the 'optimal', non-traditional, architecture generally moved to a flat (or 1-tier) command structure that had a large degree of commonality with the communication structure of the original design, with some reallocation of assets. However, when examined overall, the key finding was that teams in the 'optimal' organization did <u>not</u> elect to move to a traditional structure, and vice-versa. Teams seemed to adapt to a trigger by making (small) changes to the organization that they were familiar with, i.e., played within.

1.2 Hypothesis for Test in Experiment 3

Our working hypothesis, generated from the results of Experiment 2, was that teams, when faced with a need to change the organizational structure, would prefer to make small changes in their organization rather than large ones - even if large ones are 'best' from an objective design point of

view. Alternatively, if members of an organization are given the choice to adapt to an architecture that is 'close' (in conformity) to the one with

which they are familiar, or to one that is perceived to be quite different from their familiar one, teams would generally opt for the close design - even though the 'close' design may not have a greater performance potential to an outside observer/designer. Experiment 3 was intended to examine this issue of proximity versus optimality.

2. Overview of A2C2 Experiment 3: Elements of Experimental Design

The overarching initial hypothesis for this experiment was generated from experiment 2 and involved the trade-off between model-determined

optimality and a team's subjective feelings of comfort with their initial organization when a decision to adapt structure is being made. The activities needed to design, implement and run a model-driven, team-in-the-loop experiment to test this hypothesis, are shown in figure 1.

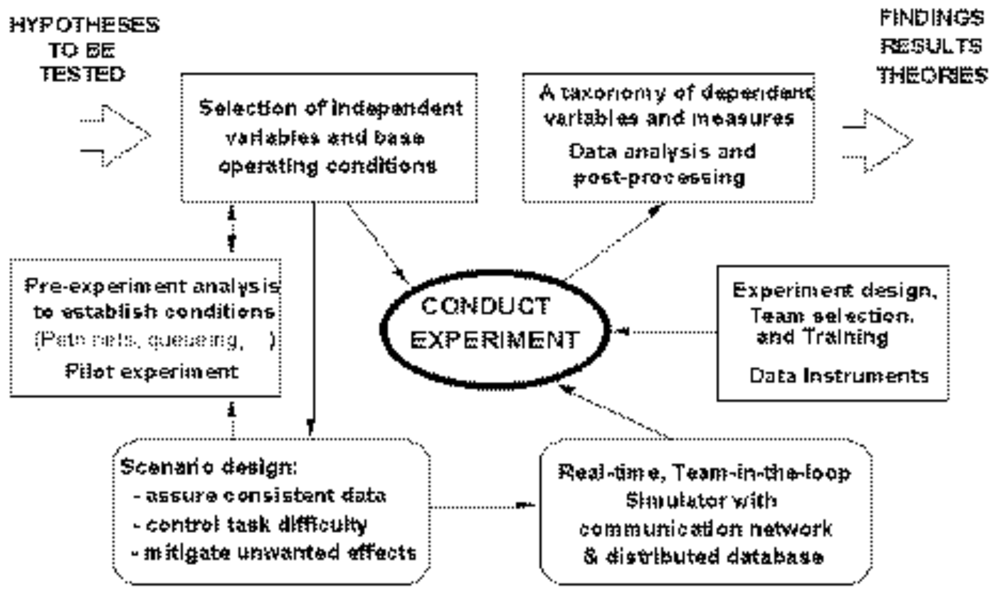
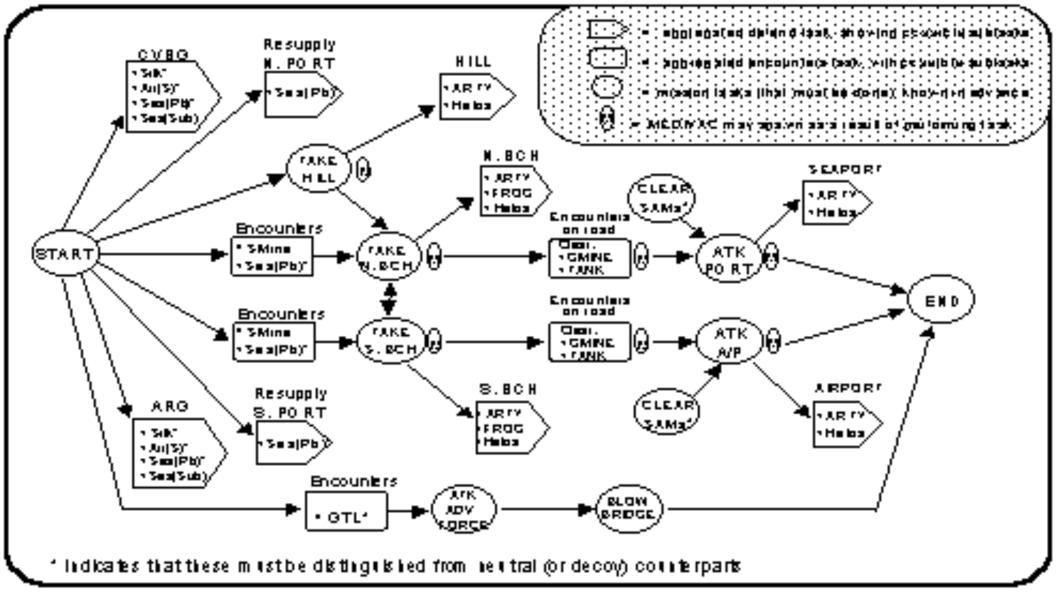


Figure 1: Elements of Experimentation with Human Teams

2.1 Scenario Design

The scenario for Experiment 3 expanded upon the amphibious mission used in Experiment 2 [Kemple, et al., 1996; Kemple, et al. 1997]. Figure 2 below details the flow of tasks for Experiment 3. Several major changes/additions were made with the intent of requiring more task coordination demands, thus increasing mission complexity.

First, we introduced a new resource for identification and precision designation (IDES), and required this capability to be used and present



when attacking certain enemy positions (such as a bridge, SAM and Silkworm sites in populated areas, etc.). Only satellite, recon aircraft (TARPS), and a SOF unit had this capability, and, as such, were required to be part of any force package assembled to attack these enemy positions. The satellite was a new asset introduced in Experiment 3 to give the scenario a flavor of future space-based capabilities. It could be "beam steered" and performed instantaneous identification of unknown enemy assets. Depending on which nodes in the organization owned the associated assets, task coordination within the team could be across several nodes, or confined to a single node if that node was designed to have the correct force (or task-organized) package. Of the three basic architectures, which consisted of a six-(initial), a five-(close), and a four-node (‘optimal’) organization, task ‘packaging’ increased as the number of nodes decreased. This meant that more inter-nodal co-

ordination was required under the initial six-node organization, but a greater number of autonomous operations were possible under the more task-organized four-node structure.

Second, a task of monitoring enemy truck movement over a pair of bridges was added. This task required the coordination of several assets. Required assets were an inserted SOF unit to detect vehicle movement, a satellite (image) to discern a missile transport-erector-launcher (TEL) from other commercial traffic, and a CAS aircraft (or a Cobra helicopter) to destroy the TEL before it could launch. A coordinated attack by CAS, engineers, and an IDES-capable platform was needed to destroy the correct bridge before the enemy could send a counterattack force over it.

Third, the team needed advanced sensor capability (satellite or TARPS) to also distinguish between enemy and decoy SAM and Silkworm sites and to distinguish enemy patrol boats from commercial shipping. Finally, the overall level of task loading on all nodes in the organization was increased by increasing the number and frequency of enemy air, sea, and ground threats. The scenario was play-tested by the lead team for ‘reasonableness’, minor modifications were made, and the scenario was provided to researchers as input to their analytical modeling activities [Curry, et al. 1997].

Having developed the baseline scenario, a 'trigger' event was developed that included the loss of approximately 30% of the available assets. The force reduction was designed so that it was still possible to perform the mission, albeit requiring a longer time. The approach envisioned for the experiment was to train teams in a baseline organization that included all the assets as in a 'mission rehearsal' mode. Following this rehearsal, the teams were given a modified operations order detailing the force reduction. They were then shown a simulated video teleconference (videotape) outlining three possible organizations that the teams could choose from. The choices were 1) A0: the baseline familiar six-node architecture with selected assets removed in a logical manner, 2) A1: an 'optimal' four-node task-organizationally designed architecture consistent with the new available asset set, and 3) A2: an intermediate five-node architecture between A0 and A1. These architectures are shown at the end of this paper in Figure 3.

2.2 **Pre-Experiment Modeling Activities**

The modelers were presented with the challenge to design a baseline organization for accomplishing the mission, given a full complement of assets similar to those in Experiment 2. This was done by UConn and NPS, using the criteria of completing the mission in minimal time while minimizing inter-node coordination of assets - yet adhering to military principles such as unity of command and unity of effort. The second modeling activity was to produce the post-trigger organizations, A1 and A2. The final design of architecture A1 settled on a *four*-person organization using the primary criteria of minimizing inter-nodal coordination of assets and a secondary criteria of minimizing workload (number of simultaneous tasks involved in) within each node. The result was an architecture that was task-organized in such a way that most tasks confronting a particular node could be performed autonomously. The design of A2 utilized a *five*-person organization using the same criteria as for A1, but in reverse order, while retaining many of the aspects of organization A0. Designed to be a compromise between A0 and A1, A2 represents an incremental shift toward the optimal, though radical, A1. Thus, A2 was designed somewhat more 'ad-hoc' than was A1. The 'distances' between the three post-trigger organizations, in each of their various structural dimensions, was examined by researchers at Carnegie-Mellon University in order to place the designs on an ordinal scale.

Having developed the three organizations, the next step in the pre-experiment design was to examine their dynamic performance, and make predictions of various individual and team performance and process measures. This is done to ensure that the numerous parameters in the scenario (task arrival rates and processing times, platform velocities, available versus required resources, etc.) are in desirable ranges to elicit the expected responses from the team. We desire a scenario that is neither too difficult nor too easy for the subjects. In particular, we desire an operating point where the salient dependent variables (to be collected) are predicted to be sensitive to the manipulated independent variables. Researchers at George Mason University conducted this pre-experiment modeling analysis with the input provided by UConn. Some iteration that modified organizational and scenario parameters was done prior to the actual experiment to fine-tune the designs.

2.3 **Conduct and Design of Experiment**

The experiment was conducted in the Systems Technology Laboratory (STL) at the Naval Postgraduate School (NPS) in Monterey. The media on which the test subjects played the scenarios was 6 SUN workstations running the Distributed Dynamic Decisionmaking III (DDD-III) software [Kleinman, et al, 1996]. The workstations had partitions between them to assure the test subjects maintained proper communications protocol, to enable information passed between nodes to be collected, and to provided some semblance of physical isolation. Communications between the test subjects were facilitated by the use of headsets at each workstation. Reflecting the future capability of a common operational picture (COP) as envisioned in Dominant Battlespace Awareness, the workstation display was identical for each node so that a common picture was available to all team members. Assets owned by an individual team member (node) were uniquely color-coded by position/node to help ease identification and coordination.

Prior to conducting any DDD runs, each team had to undergo adequate training to be able to operate in the simulation environment. All subjects were given one hour of "buttonology" training to familiarize them with the DDD interface. Before the first trial, each team was given one hour of team training to get used to the individual station they would be playing and familiarize themselves with the communications procedures. The teams were scheduled for two 2-hour DDD trial runs for data collection. A two-hour planning session was conducted between the first and second runs in order to collect data on team decisions as a result of the trigger event.

The experiment was designed with nine teams first being trained in organization A0, with full assets on a scenario similar, but not identical, to the one they would initially face in the actual experiment. Test subjects were provided with background information consisting of an Operations Order, which explained the scenario, mission, friendly and enemy assets, and execution instructions. Handouts were provided to each team member before each trial run showing the mission priorities, friendly asset starting positions, and a list of tasks requiring coordination among the players. These handouts were designed to help players adjust to different command structures with minimal learning. Test subjects were provided with a DDD tutorial prior to the first run in order to help familiarize them with the DDD simulation interface. Following the first run, but prior to the first data collection run, the teams entered a 2-hour planning session where they were presented with the modified operations order, the new (reduced) asset set, and with a choice of playing in one of three organizational structures: A0, A1, and A2. The two post-trigger runs were counterbalanced between the teams' choice and either architecture A1 or A2 (experimenter specified). This was done to insure that each architecture was run a number of times (none was omitted) as well as to avoid any effect of team learning as they played the scenarios. After each run, the team entered an after-action review session. The hope was that over the 18 data collection runs, we would obtain (through choice and specification) approximately six runs in each of the three organizations for subsequent data analysis.

Subjects in the experiment consisted of 9*6 = 54 postgraduate students drawn from two courses at NPS. The subjects were assigned to teams based on their schedule of availability and degree of operational experience. Prior to the experiment, researchers at Alphatech, Aptima, and NPS developed data collection forms and instruments that were designed to capture the processes behind the teams' choice of architecture, and the salient elements that entered their choice. The lead team of NPS students served as observers and data collectors for this process.

During the experimental runs, data collection included: (1) automated data logging and measures collection via the DDD computer software, (2) observer logging and categorization of voice communications, and (3) lead-team (subject matter experts) provided ratings on subject performance and workload. Post experiment data collection in after-action reviews solicited (via instruments and questionnaires), other information on team processes and subjective opinions.

For data collection, each trigger planning session and trial was video taped. A separate cassette player was also used to record pretrial planning, the trial, and the after-action review conducted by each team following the trial run. The video and audio tapes provided data in their own right and served as a backup to manual survey-style data collection.

2.4 Enhancements to the DDD Simulator

The previous experiments were conducted using the Distributed Dynamic Decisionmaking (DDD-III) software tool [Kleinman, et al., 1996]. This software is a distributed team-in-the-loop simulator for conducting controlled laboratory-based experiments in a real-time environment. In the course of the A2C2 project, the DDD has been enhanced to treat Joint military environments, and to allow manipulation and control of the key dimensions of both mission/task and organizational structures. For experiment 3 additional modifications were needed to treat the new features added to the scenario (high inter-node coordination and classification of contacts as either hostile or neutral /decoy) and to capture and log new elements of team response. In addition, suggestions for improvements made by subjects after Experiment 2 were made to enhance ease of interface use and reduce training time.

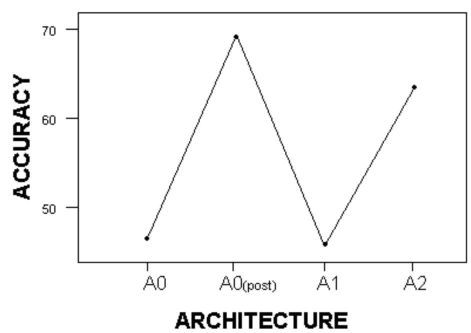
3. Findings and Results

The primary finding, with respect to team adaptation, was unexpected. All nine teams, when presented with the reduced assets mission, elected to stay within their original A0 organization - despite the fact that *a priori* model analysis predicted that this organization would exhibit poorer performance than either A1 or A2. Consequently, in the subsequent data collection runs to examine actual performance we had nine runs with teams playing in A0, but only four or five runs with teams playing in A1 and A2, respectively. The analyses of the experimental data (including the planning session data and after-action reviews) are currently ongoing by all of the A2C2 research groups. Of particular relevance are the comparisons of team performance and process measures across organizations, and with model-generated predictions. Initial indications suggest that the teams performed better in A0 than in A1 or A2! If such is the case, model assumptions and design criteria will need to be re-evaluated, and the model-test-model paradigm exercised again for yet another experiment.

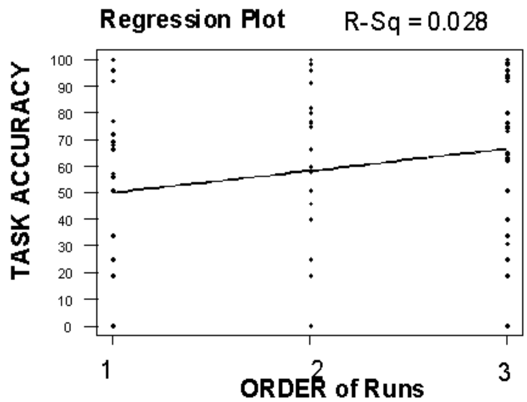
3.1 Lead Team Results

In support of Experiment 3, the Lead Team performed many roles such as scenario development, team training, handout generation, laboratory preparation, scenario play-testing, and data collection. In addition to collecting data for the A2C2 research team, the Lead Team, to satisfy requirements for a Systems Evaluation class, determined unique measures for data collection and analysis. The Lead Team was interested in the relationship between inter-nodal coordination and task performance. An hypothesis of the architectural design process is that an organization designed to minimize inter-nodal coordination would also perform better. The reasoning is that such an organization would operate under a lower workload. (We assume that inter-node coordination is the major contributor to workload at a node.) In addition to performance, the Lead Team decided to analyze whether learning continued as the teams performed more runs. This was driven by observations during the runs indicating that some players lacked the basic concepts/skills of performing coordinated attacks or of matching required with available resources, etc.

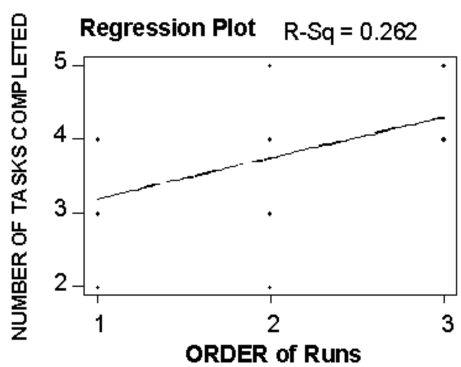
A complete analysis of the Lead Team's hypotheses is contained in a recent NPS thesis [Benson, 1998]. Results of the performance hypothesis showed that the post-trigger A0 organization had the highest average task accuracy, while the A1 organization had the lowest average accuracy (see figure 4). This was a major finding in that it contradicted the model-based (predicted) performance trends. It has been suggested that one reason for this result may have been inadequate training of the subjects to perform coordinated attacks (i.e., with multiple assets) at a single node. These attacks would have been more prevalent in A1 as this organization has only four players to complete the required mission tasks. Attacks requiring multiple assets under the A0 or A2 organizations could more easily have been performed by more than one node's participation, which was perceived to be easier than an autonomous attack. The result was a noteworthy decrease in average task accuracy in the A1 architecture.



Based on findings from the analysis of the performance measures, the Lead Team determined that the "learning effect" should be examined. The final hypothesis examined by the Lead Team was derived from the possibility that subject teams continued to improve performance throughout the sequence of runs. The basic assumption that each subject team had a solid understanding of the DDD interface and how to accomplish the assigned tasks required of each decision maker was a topic of particular interest to the Lead Team. We hoped to determine whether teams continued to improve as they completed more runs. Linear regression analysis produced some interesting results. Regression was done on the accuracy versus the order in which teams did runs (see figure 5). The regression resulted in a slight but significant upward slope ($p = .051$). The positive slope suggests that accuracy increased as the teams completed more runs.



Additionally, the number of tasks completed was examined. The result was an increase in the number of tasks completed throughout subsequent runs (see figure 6). Teams accomplished more tasks in the later runs than in the first run regardless of the architecture they were operating under.



This analysis suggests that learning was in fact occurring throughout the experiment. When the tasks not completed were assigned an accuracy of zero and were included in the analysis, the accuracy increased from run to run, and when they were omitted, the number of tasks accomplished increased from run to run without loss of accuracy. This is consistent with the belief of the Lead Team observers that player proficiency was lacking into the post-trigger runs. In some post-experiment responses, players stated that their performance increased as they did more runs and became more familiar with the interface and scenario. Adequate training should, in the future, significantly reduce this effect.

4. **Follow-up Experiment**

Due to the discrepancy between the analytical predictions of performance and the subsequent results, it has been suggested that a follow-up experiment to Experiment 3 should take place. The goal of such an experiment would be to explore the possible reasons for the difference between expected and actual performance of the organizations. Sufficient team and individual training would be paramount to avoid a repetition of the possible learning effect that occurred during Experiment 3. It has been proposed that this interim experiment should take place during the summer quarter at NPS. Experiment questions and hypotheses are under development at this time and should yield confirmation of some of the suggested results of Experiment 3.

5**. References**

Benson, R., "Conduct and Assessment of A2C2 Experiment 3" Masters Thesis, Naval Postgraduate School, Monterey, CA, June 1998.

Curry, M.L., D.L. Kleinman, and K.R. Pattipati, "Mission Modeling as a Driver for the Design and Analysis of Organizations," *Proc. 1997 Symposium on Command & Control Research*, Washington, DC, June 17-20, 1997.

Kemple, W.G., D.L. Kleinman and M.C. Berigan, "A2C2 Experiment: Adaptation of the Joint Scenario and Formalization," *Proc. 1996 Symposium on Command and Control Research and Technology*, NPS, Monterey, CA, June 1996.

Kemple, W.G., J. Drake and D.L. Kleinman, "Experimental Evaluation of Alternate and Adaptive Architectures in Command and Control," *Proc. 1997 Symposium on Command & Control Research* , Washington, DC, June 17-20, 1997

Kleinman, D.L., P.W. Young and G. Higgins, "The DDD-III: A Tool for Empirical Research in Adaptive Organizations," *Proc. 1996 Symposium on Command and Control Research and Technology*, NPS, Monterey, CA, June 1996.

Y.N. Levchuck, K.R. Pattipati and M.L. Curry, "Normative Design of Organizations to Solve a Complex Mission: Theory and Algorithms," *Proc. 1997 Symposium on Command & Control Research* , Washington, DC, June 17-20, 1997. [Best Paper Award]

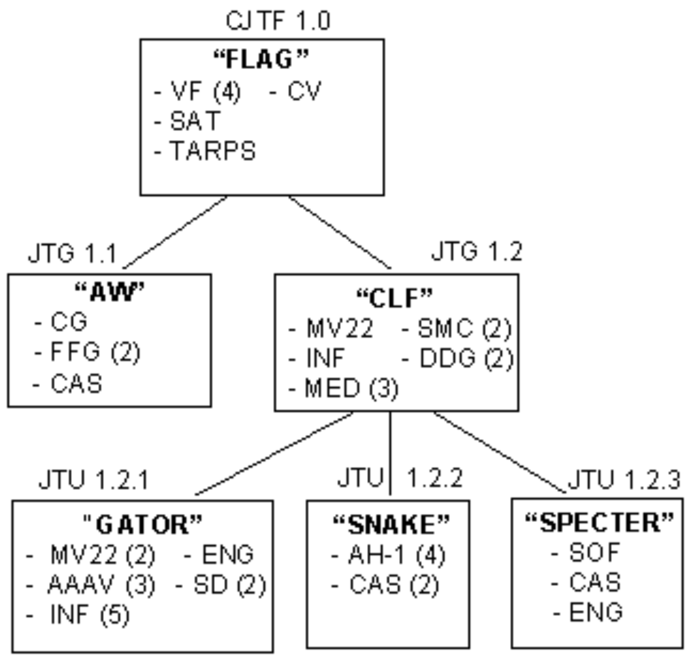


Fig. 3a: Architecture A0 (pre-trigger)

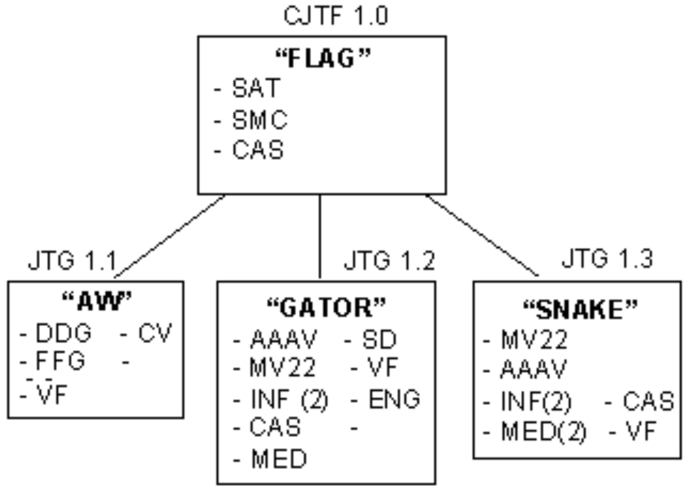


Fig. 3c: Architecture A1 (post-trigger)

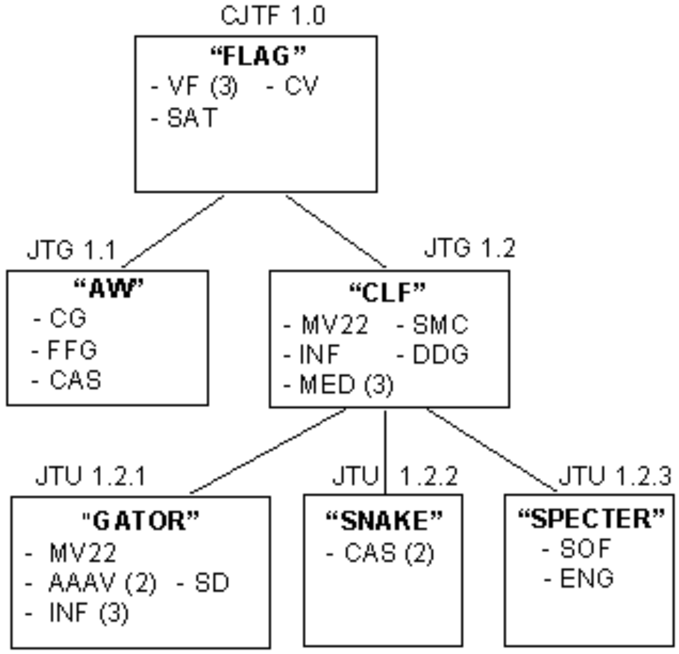


Fig. 3b: Architecture A0 (post-trigger)

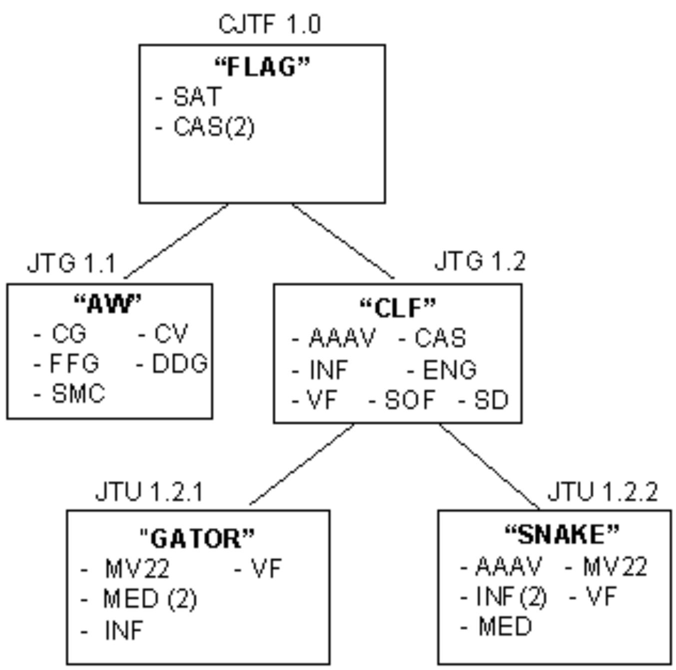


Fig. 3d: Architecture A2 (post-trigger)